Chapter 14

Interferometer Design for Synthesis Imaging

David Mozurkewich

Naval Research Laboratory
Washington, DC

This Chapter reviews the problem of synthesis imaging at optical and infrared wavelengths, how to design interferometric arrays, and how to build the beam combination optics.

14.1 Good Fringes vs Good Science

The present generation of interferometers has a sensitivity limitation that restricts us to observing bright stars. In the future, the Keck interferometer and other arrays may exceed this limit, but until then we are restricted to observing bright stars.

Since we are using instruments that can only observe stars, we should admit up front that we are observing stars and we must assert with great confidence that stars are interesting. For this to be true, we need to make images of stellar surfaces. With these images we can see spots form and evolve and we can watch the stars rotate. Perhaps we will see the spots form at preferred latitudes and perhaps those preferred latitudes will change in a cyclic way as they do on our Sun. Stars are interesting with these kinds of observations, and so we need to make images.

A large portion of this talk will cover imaging strategies. That may seem like an odd topic. After all radio astronomers taught us how to make images 40 years ago. But their images are at longer wavelengths. For visible light observations, they only solved half the problem. To see what is missing, look at Figure 14.1. It shows the visibility amplitude

versus projected baseline length for two stars, one with a featureless uniform disk and one that is almost as featureless but with limb-darkening. What is most obvious is that the two visibility curves are nearly identical on short baselines.

The situation is similar in Figure 14.2, which shows one of the interesting stars I want to image along with one of the boring stars from Figure 14.1. The interesting star has limb-darkening *and* a spot. Again, on short baselines the curves are very similar. If I were to change the diameter of this spotted star a little bit, the position of the first minimum would shift and we would have virtually indistinguishable curves out to the first zero. On longer baselines there is a factor of two (or more) difference in the visibility amplitude between the stars, but these amplitudes are low—less than 0.1.

To summarize, if we want to make high dynamic range images of stars, or even separate stars with any surface structure from those without, we have to measure low visibility fringes.

Fringes are easiest to detect on short baselines, but the interesting science requires the long baselines. It can be extremely difficult even to find fringes on long baselines. Another complication is that we have to track the fringes. The atmosphere moves the fringes around by a lot more than a few wavelengths, and on the long baselines where fringes are weak there is not enough signal-to-noise in a short integration* time to find them and position the delay lines properly. A long integration time does not work because the fringes are washed out by the atmosphere.

There is the problem. Stated somewhat differently, you have a choice of either recording useful data *or* looking at interesting sources. This is not the kind of choice I like to make.

The existing optical interferometers do not have a lot of dynamic range and their sensitivity drops dramatically with fringe visibility. The signal-to-noise is a monotonically increasing function of NV^2, where N is the number of photons detected in an integration and V is the fringe visibility amplitude. To observe a star on a baseline where its visibility amplitude is 0.1 requires 100 times as much sensitivity as an observation of the same star on a baseline where it is unresolved. The low visibility amplitude observation needs 100 times as many photons in an integration to give the same signal-to-noise. This is a loss of 5 magnitudes, and 5 magnitudes is a lot of sensitivity to lose when the limiting magnitude for unresolved stars is in the range of 5 to 8. We absolutely need have a scheme to get around this sensitivity limit if we are going make useful images. Consider the following five possibilities.

*In this Chapter, I use integration to refer to data that goes into a single estimate of the fringe parameters. An observation is a collection of integrations that are combined to increase the signal-to-noise to a useful level.

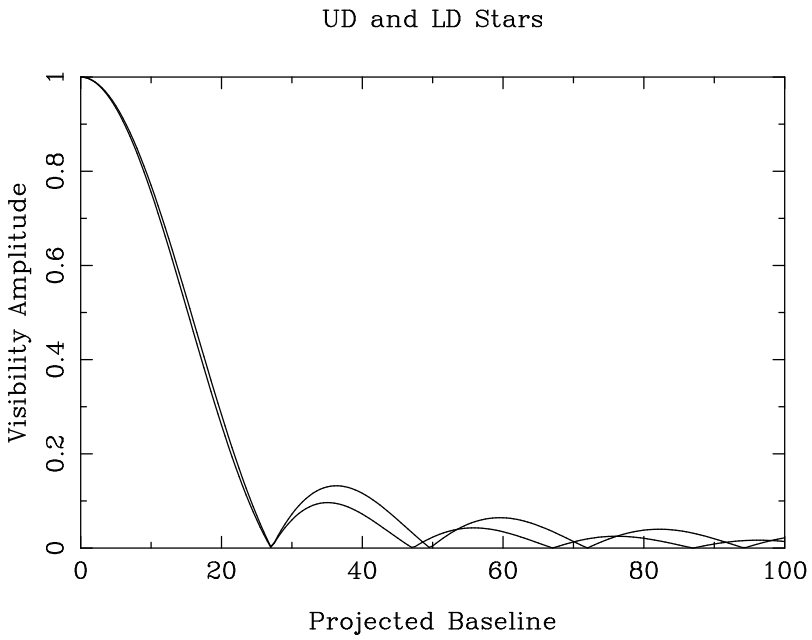


Figure 14.1: Visibility amplitude as a function of projected baseline for uniform and limb-darkened stars.

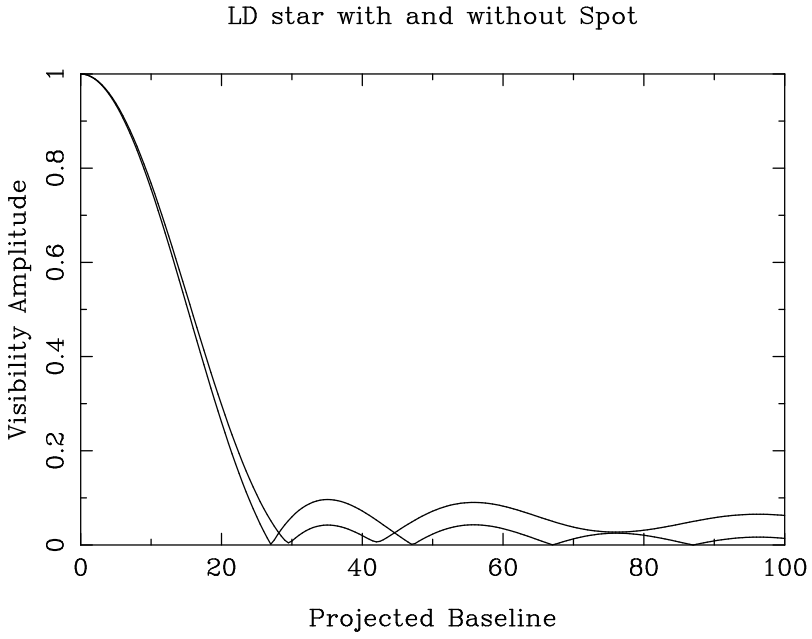


Figure 14.2: Visibility amplitude as a function of projected baseline for the same diameter: a limb-darkened star with and without a central spot.

14.1.1 Alternative 1: Integrate Forever

The simplest scheme to try is to make up for the lower signal-to-noise by observing longer. This does not work. A star with a visibility amplitude of 0.1 requires 100 times as many photons per integration. But we cannot integrate 100 times as long to get those photons because the maximum integration time is limited by phase fluctuations in the atmosphere. We have to use the data from short integrations to estimate the visibility amplitude (or rather the square of the amplitude). This will be a noisy estimate. The signal-to-noise can be improved by averaging a large number of estimates, and will increase by the square root of the number of integrations. The signal-to-noise ratio is given by

$$\mathrm{SNR} = \frac{V^2}{\sigma(V^2)} = \frac{2}{\pi^2}\frac{NV^2}{\sqrt{1+\frac{4}{\pi^2}NV^2}}. \tag{14.1}$$

This is a monotonically increasing function of NV^2. In the high signal-to-noise regime, $NV^2 > 3$ per integration, the required observing time increases as V^{-2}. A star whose visibility amplitude is 0.1 requires 100 times the integration time. This converts tens of seconds into twenties of minutes. A reasonable observation time. But most of the objects we want to image will be in the photon-starved regime, $NV^2 < 2$ per integration. Here, the integration time increases as V^{-4}; a 1-second observation is converted into 3 hours. However bad this may seem, the actual situation will be worse, since this is an upper bound to how well we can do. In the photon-starved regime, we cannot determine the phase of the fringe—or even if it is present—during the atmospheric coherence time. Hence, we cannot fringe-track and we have to blindly point the delay lines. We are required to build an interferometer that allows us to control the delay *open loop* to better than the coherence length of the starlight. But the atmosphere is moving the fringe delay around by typically 50 μm peak-to-peak. This requires a coherence length longer than 50 μm and we need to restrict the fractional bandpass to something on the order of 1%. To observe a low visibility amplitude fringe, we need to restrict the bandpass and hence the number of photons detected per integration. NV^2 is reduced yet again and the single integration signal-to-noise drops even more.

Another way of blind pointing—running an interferometer without fringe tracking—is to use a broad-band channel and scan in delay across the entire region where the fringe could occur. The power spectrum of these data will show a peak whose shape is proportional to the shape of the optical bandpass and whose amplitude is proportional to N^2V^2. In this approach, we do not lose due to a narrow bandpass, but a large fraction of the integration time is spent at delays where there are no fringes. This adds noise to the data and is little better than the narrow-band approach.[†] In addition, the broad-band visibility is harder to interpret. My conclusion is that there is no really good way to just integrate forever in the hope of seeing a fringe.

[†]A third variant is to combine these methods using a spectrograph to simultaneously record 100 channels of narrow-band data. Here, detector read noise is the limitation.

It is true that for very short baselines, say a few meters, the atmospheric fluctuations are small enough that integrating forever can produce results. However, long baselines are needed for imaging stars, so this is not an option for the problem which I am interested in.

14.1.2 Alternative 2: Sources with High Visibility

A second alternative is to restrict ourselves to looking only at sources that provide plenty of signal for fringe tracking. There are indeed sources that have high visibility at the same baselines where there's structure, as shown by the following examples:

Figure 14.3 shows a binary star. The visibility varies wildly as function of baseline length. Even if you can only observe on the baselines where the visibility amplitude is high, you can measure the period and amplitude. When you measure how fast the envelope falls off with projected baseline length, you estimate the diameter of the primary star, and if you measure how the mean amplitude drops, you also get the diameter of the secondary. A lot of information can be obtained from this type of observation. There are people, such as Christian Hummel at the US Naval Observatory, who are making a little cottage industry out of observing double stars with optical interferometry, and they are doing some great science.

Figure 14.4 shows a star with a uniformly-bright, circumstellar disk. This is a very nice source to observe. What we see here is the fringe pattern of the disk superimposed on the fringe pattern of the star. The drop in visibility amplitude at short baselines occurs when we resolve the size of the disk. The long plateau is formed by fringes from the star. The wiggles superimposed on the plateau are from the disk. Their size and location tell us about structure on the disk for spatial scales larger than the star. The star embedded in the disk has converted our problem from one of detecting low visibility fringes to one of measuring small variations in higher-amplitude fringes.

There are sources like this. At infrared wavelengths, from 2 to 10 μm, a number of young stars have disks and old stars have shells. There are enough of these sources to do a fair amount of science. However at visible wavelengths, you have only Be stars, and perhaps μ Cep, but I'm not sure there's anything else that fits into this category and is also bright enough for the current generation of interferometers.

Unfortunately, with this alternative you are letting the instrument tell you what science you are allowed to find interesting, whereas, if possible, you should be building an instrument that will let you observe the sources that interest you. For this reason, I do not like Alternative 2—even though it is a valid imaging strategy.

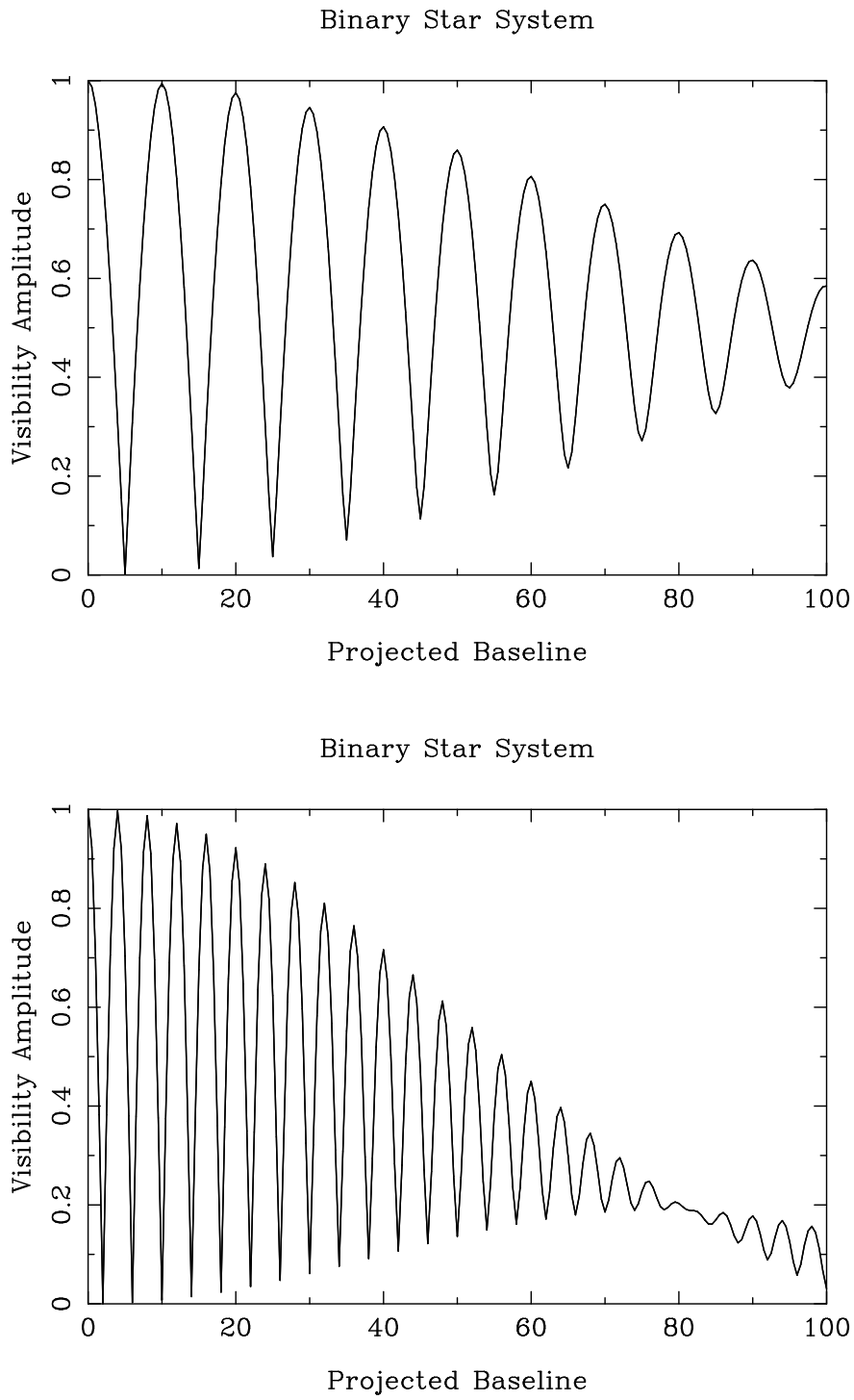


Figure 14.3: Visibility amplitude as a function of projected baseline for two binary star systems with different stellar diameters. The system shown in the upper plot has equal brightness components.

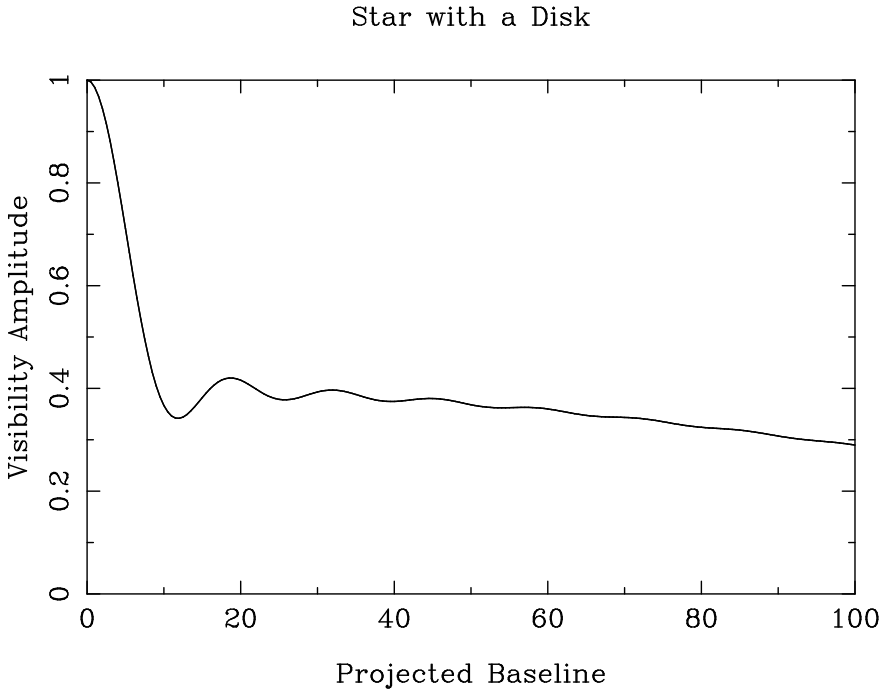


Figure 14.4: Visibility amplitude as a function of projected baseline for a compact star surrounded by an extended disk.

14.1.3 Alternative 3: Wavelength Bootstrapping

We have been looking at plots of visibility versus projected baseline. This is not the whole story. The ordinate should be spatial frequency. The spatial frequency is the projected baseline divided by the observation wavelength. So far we have only discussed what happens when we change the baseline between the telescopes. We can also change the wavelength. A broad-band (multi-wavelength) system can simultaneously span a significant range of spatial frequencies on one of these curves. The longest wavelengths will usually have the highest visibility amplitudes. The shortest wavelengths always have the most resolution. The fringe tracker uses the highest-visibility wavelengths while the science detectors record the low visibilities. This is referred to as "wavelength bootstrapping." Figure 14.5 shows an example from the Mark III. The star is Arcturus. Fringe tracking was done in the 700–800 nm range while simultaneously, observing at 550 nm. Because the visibility amplitude is a very steep function of baseline length, there is enough visibility at the longer wavelength (i.e., shorter "baseline") to determine the fringe position and stabilize it. There are measured squared visibility amplitudes in this figure almost as low as 10^{-4}. That's low, corresponding to 5 magnitudes of sensitivity loss in the photon-rich regime and up to 10 magnitudes in the photon-starved regime.

The main advantage of wavelength bootstrapping is that the fringe tracking data can be used to stabilize the fringe—either in hardware or software. With this accomplished, it

is possible to overcome the atmospherically induced limit on the integration time. This allows us to push the data from the photon-starved regime into the photon-rich regime and can shorten the observation time needed by a factor of 100 or more. Note that it is not necessary for the star to be the same shape at all the wavelengths. It is only necessary for it to be compact at some wavelengths and resolved at others.

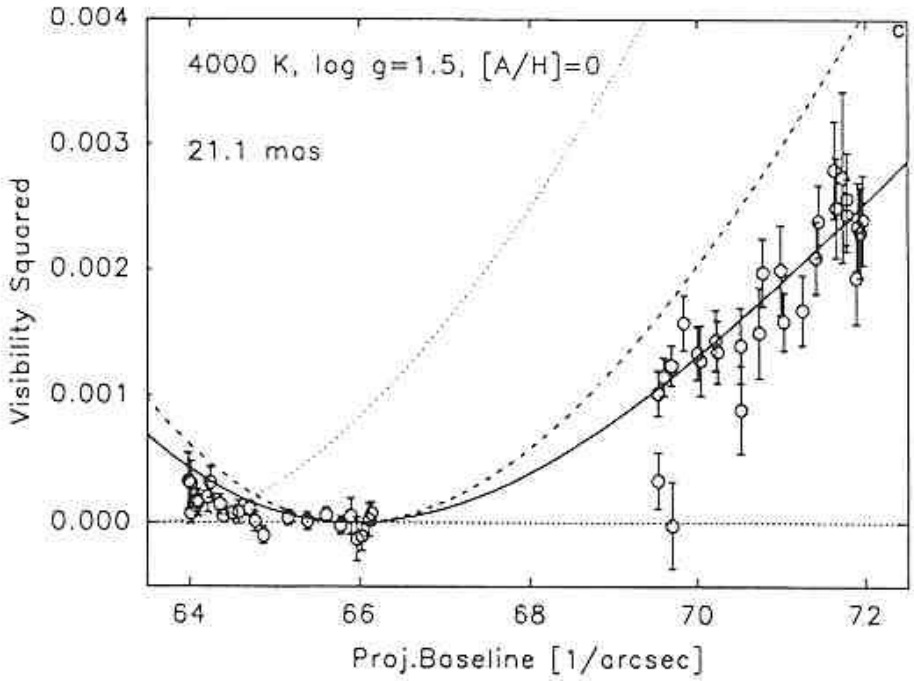


Figure 14.5: Squared visibility amplitude versus projected baseline for Arcturus showing the application of wavelength bootstrapping. This 550 nm data was taken while fringe tracking in the 700 to 800 nm band (Quirrenbach *et al.*, 1996).

Although wavelength bootstrapping is useful, it has limitations. The fringe tracking is performed at the longest wavelength, but it must be done with enough accuracy and speed to stabilize fringes at the shortest wavelength. Also, unless we have good adaptive optics, we must limit the size of the aperture to that which is allowed at the shortest wavelength since the atmosphere introduces a relative piston between two apertures that are co-located but of different diameters. These limitations on the aperture size and integration time severely limit the sensitivity of the long wavelength measurement. And even if we have enough sensitivity for the observations, there is still the problem of knowing the index of refraction of air with sufficient accuracy (due to the changing and unknown water vapor content) to predict the fringe position from a measurement at a greatly different wavelength. Because of these types of effects, it will be difficult to apply wavelength bootstrapping over more than a factor of two or three in wavelength. But a factor of two or three in wavelength corresponds to a factor of two or three in resolution and that can make a big difference in the type of science that can be done. I dislike interferometer designs where the blue light is siphoned off for use in angle tracking and adaptive optics leaving only the long wavelength

portion of the bandpass for fringe measurements. I think wavelength bootstrapping *must* be an integral part of any imaging interferometer.

14.1.4 Alternative 4: Baseline Bootstrapping

It is also possible to stabilize the fringes on a long baseline through "baseline bootstrapping." Figure 14.6 shows a 5 telescope linear array. The observing procedure is to simultaneously fringe track on baseline A–B, and on baseline B–C, and on baseline C–D, etc. Although only five are shown, you may string together as many telescopes as you want. The change in phase on each of the short baselines is determined, and those phases are combined in various ways to obtain the change in phase on each of the long baselines.

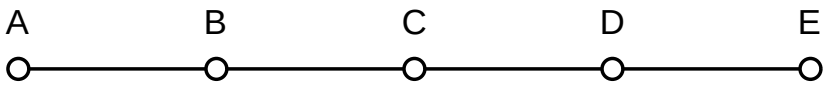


Figure 14.6: A redundant linear array useful for baseline bootstrapping.

This procedure works because of the closure-phase relationships. The change in delay due to the atmosphere and the instrument affects all the light intercepted by one telescope in the same way. Thus the phase changes induced on the path from the star through telescope "B" to the beam combiner adds exactly as much phase to baseline A–B as it subtracts from baseline B–C. Changes in the baseline phase due to source structure do not cancel out in this procedure and that is good because those are the phases we want to measure. But these phases vary slowly—at least a lot more slowly than the atmosphere. Although source phases changing with time limit the maximum integration time in principal, they do not impact the usefulness of this method.

The goal here is the same as it is with wavelength bootstrapping; estimate the phase on the longest baseline with enough accuracy to stabilize it. Then the integration time can be extended beyond the atmospheric limit, and in the process a couple factors of V^2 are removed from the required observing time. But this method is not perfect: When N short baselines are strung together to phase a long baseline, the phase noise on all the short baselines contributes to the uncertainty in the long baseline phase. The phase noise on a long baseline will be $\sqrt{N}$ times the phase noise on a short baselines. As a result, the system visibility is lower on the longer baselines. But as long as we are aware of this effect and calibrate, it should not be a limitation.

Figure 14.7 shows an example of baseline bootstrapping data from three baselines of the NPOI. The squared visibility amplitudes for the two short baselines are shown in the upper panel. The long baseline data is shown in the lower panel. The signal-to-noise on the longest baseline is too low for the fringes to be detected or tracked during the observations.

Nevertheless, the data was obtained because the fringes were detected and tracked on the two short baselines. This technique is just starting to be exploited and is a very good way to make an imaging interferometer work.

14.1.5 Alternative 5: Guide Star Methods

The fifth scheme is the use of a guide star. If the fringes are too faint to allow fringe tracking on the star itself, you need a method to estimate the delay, or at least the change in delay, other than by observing the fringes on that star. The obvious technique is to use the delay measured on a nearby star. In radio interferometry, this is standard operating procedure and is used to remove phase variations induced by the ionosphere; the telescopes switch back and forth between two sources. The phase variations observed on the bright source are applied to the faint source, allowing longer integration times.

At optical wavelengths, the atmosphere varies with a coherence time on the order of 10 ms. The two stars must be observed simultaneously since this is too fast to allow switching between the stars. Fringes are only tracked on the guide star but the delay corrections deduced from that tracking are applied to both stars. The phase, however, decorrelates with increasing angle as well as time. At visible wavelengths, the maximum allowed angle at the best astronomical sites is less than ten arc-seconds. 10 arcsec regions surrounding the brightest stars only covers a very small fraction of the sky. This does not make for a practical instrument.

In the infrared the atmosphere is more benign; the pathlength errors it introduces are a smaller fraction of a wavelength and do not hurt the observations as much as they do at shorter wavelengths. This technique is potentially very exciting for observations at 10 μm. At shorter wavelengths around 2 μm the system visibility amplitudes will be low, but the technique will certainly produce some good science. Calibration at these wavelengths is going to be interesting: The visibility amplitude has already been significantly reduced by decorrelation between the two paths, and so small changes in the seeing will cause small changes in that decorrelation which can result in large changes in the calibration. The system visibility amplitude may be such a strong function of the seeing that it will be difficult or impossible to calibrate the observations. Recently, I heard a rumor that measurements from the Palomar Testbed Interferometer suggest I am wrong. I hope the rumour is true, and I am anxiously awaiting the publication of these results.

I should point out that guide star methods are probably restricted to natural guide stars. Laser guide stars that light up the atmosphere are intrinsically incoherent, and as a result, you can use them to correct the wavefront flatness but not absolute phase.

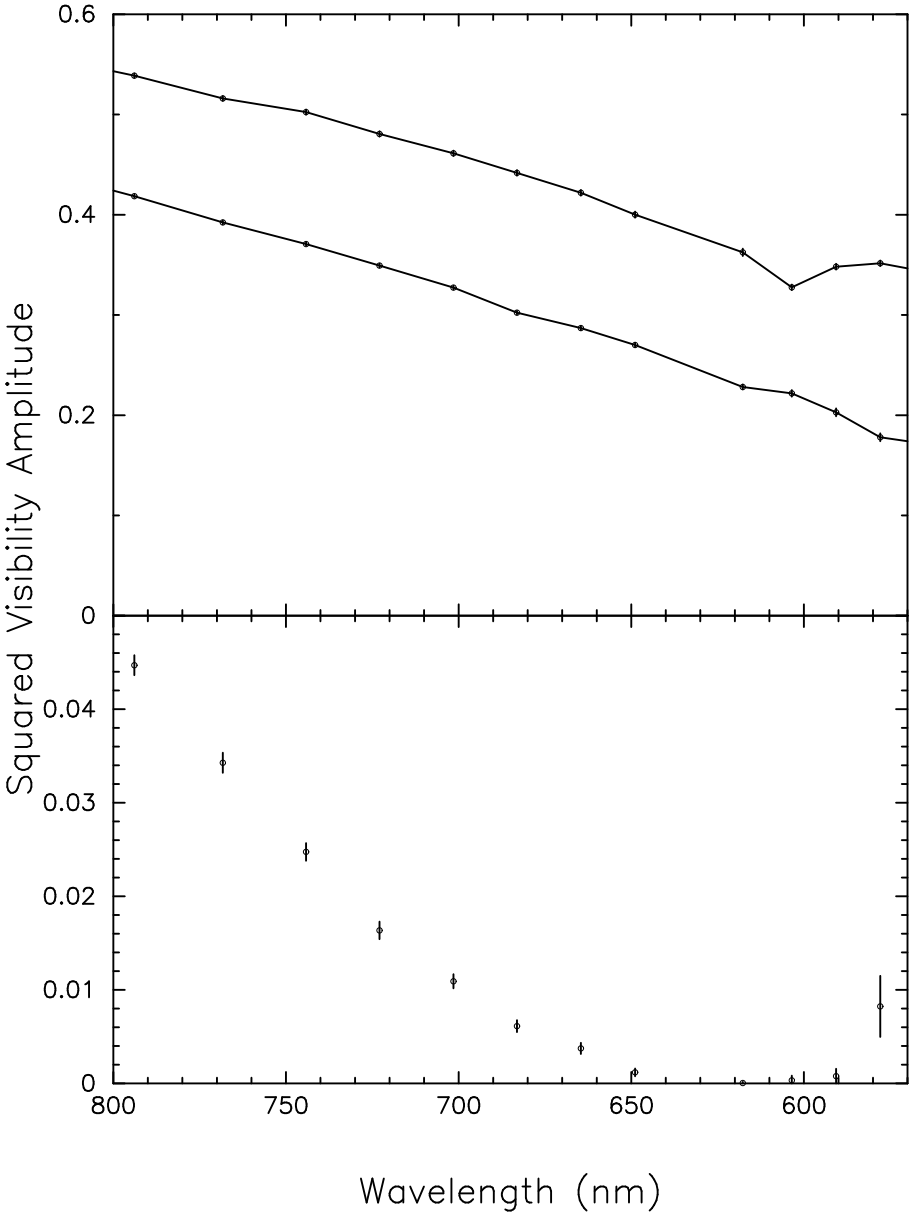


Figure 14.7: Fringe data as a function of wavelength from the NPOI. The data was taken simultaneously on all three baselines. The longest baseline (lower curve) was only obtained because the fringes were stabilized on that baseline using the two shorter baselines (upper curves).

14.2 Design of Interferometric Arrays

14.2.1 Optimal Design of Interferometric Arrays

A lot of work has gone into how to design arrays and every few years someone will come by and say "This is the right way to do it." There are plenty of examples. I am not going to spend much time discussing them because most of this work is not relevant to our problem.

- **VLA Y.** The VLA was laid out in a Y-configuration with a power law spacing of the elements along the arms. This gives good images using Earth-rotation synthesis and is often quoted as the correct way to layout an array (see for example Mathur 1969 and Chow 1972).
- **Random Arrays.** When the telescopes are arranged in a geometric pattern such as a **Y** or a circle, some of that pattern bleeds into the (u, v) coverage with the sample points falling along lines and arcs. This patterning can be removed by laying out an array with a more random distribution. The actual configurations are determined by trial and error or exhaustive searches and give more uniform (u, v) distributions that lack the patterning seen in more conventional arrays.
- **Special Purpose Arrays.** There is an interesting masters thesis from MIT recently (see Kong *et al.* 1998) where the author looked at how to arrange the apertures to create regions in the image close to the central peak that are devoid of sidelobes. His application was to design an array that is optimized for a planet search. Other optimizations should be possible for whatever applications are of interest.

There are a lot of possibilities (Mathur, 1969; Golay, 1970; Chow, 1972; Cornwell, 1988; Keto, 1997). I doubt very much if the "optimum" designs are truly applicable to *synthesis imaging at optical wavelengths from the ground with Michelson interferometers.* This is because, although we can argue about the existence of a *best* array configuration, there already exist a lot of *pretty good* configurations. And it turns out the pretty good configurations are not that much worse than the best configuration, whatever that is. As a result, practical considerations will win out over the optimization processes involved in these array designs.

14.2.2 Partially Redundant Arrays

In my view, imaging requires redundant arrays. If we are going to image fine-scale structure on the surface of stars, we need to measure visibilities on baselines where the fringe visibility is almost unmeasurable. This requires baseline bootstrapping. The best arrays for baseline bootstrapping are redundant arrays. Arrays that are redundant enough to allow baseline bootstrapping can have remarkably good (u, v) coverage. I say partially redundant because a completely redundant array has all the elements evenly spaced along a line. You do not want a completely redundant array.

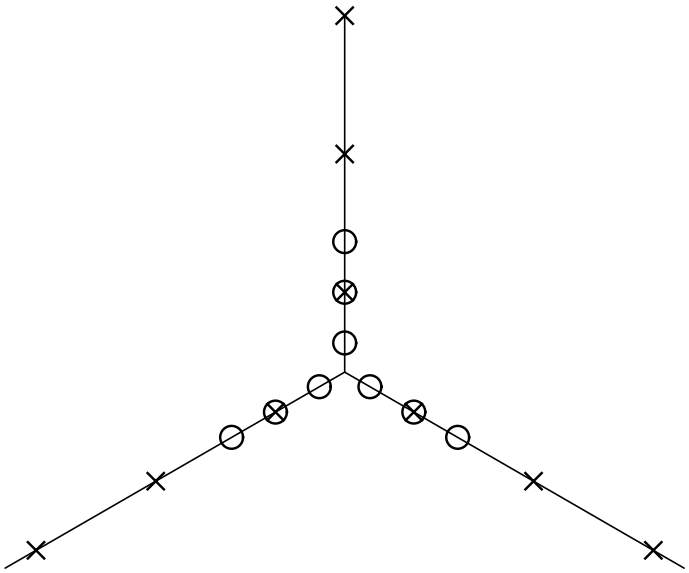


Figure 14.8: Layout of the NPOI showing its partially redundant configuration.

14.2.3 Array Design with Vacuum Feed Systems

It seems obvious to me that imaging at optical wavelengths requires a vacuum feed system. The star light has to be transferred from the telescopes to the optics lab and there is just too much air along that path. That path is horizontal and close to the ground. There is always turbulence along the ground. If the path is not evacuated, there will be problems with local seeing along that path. If the path is not evacuated, there will be problems with dispersion because the telescopes are not all at the same distance from the lab and after adding an air delay there will be more air in one path than in the others. There can be problems with refraction because simple measures to control seeing will cause the air to stratify. Just to get rid of these problems the feed system should be in a vacuum pipe. If the feed system is in a vacuum, then a random array, or even a circle, will have pipes strewn all over the site. You will want to collect the pipes together into lines or a Y for convenience. This is particularly true when you realize the telescopes have to be moved to match the resolution of the array to the size of the source being imaged. It would be nice to reuse the same pipes and telescope stations for different array layouts. This leaves us with a partially redundant Y. For purely practical reasons this seems to be the only reasonable choice and is probably the way we are going to build arrays.

14.2.4 NPOI Configuration

The NPOI is a good example of what I mean by a partially redundant array. The layout of the inner part of the NPOI is shown in Figure 14.8. Two array configurations are shown; one marked with circles, the other with an $\times$. In the complete system, all stations are part of two configurations. The nearest neighbors within a configuration are all equally spaced

allowing baseline bootstrapping. For a six-telescope configuration with three telescopes per arm, the longest baseline is $(2\sqrt{3}+1) = 4.464$ times longer than the short baselines. The six-element completely redundant (linear) array has a maximum baseline five times the shortest, roughly 10% longer than the NPOI design. A non-redundant six-element array consists of 15 unique baselines. In the partially redundant NPOI design only two baselines are redundant, 13 are unique. By contrast, the fully redundant, linear array has only five unique baselines. This partially redundant type of design seems to be an excellent compromise between (u, v) plane coverage and baseline bootstrapping.

The (u, v) plane coverage obtained from this six-telescope design is asymmetric with highest resolution perpendicular to the unpopulated baseline. For best imaging, we need to observe the same target three times, once with each arm unpopulated. Or we need to have nine telescopes.

14.2.5 Polarization and Beam Rotation

When designing an interferometer, you must remember that the optics are not idealized entities. It is important to specify, or at least understand, all optical properties of each element in the system. Arguably the hardest part of building an interferometer is providing specifications for all the components. For example, consider the simplest optical component in the system, a mirror. For building a telescope, all we have to do is know the surface quality and reflectivity.

For an interferometer, the specifications of a mirror are more complicated. A root-mean-squared surface quality is not enough since the micro-roughness can dominate the far-field reflectivity. The reflectivity is usually given for unpolarized light at normal incidence. This is good enough for a telescope where most of the reflections are near normal incidence, but for an interferometer, the polarization dependence is important. Using the unpolarized reflectivity will provide estimates of the system throughput that can either greatly underestimate or overestimate performance. The sign of the error depends on the orientation of the mirrors in the optical train.

Mirror coatings not only absorb and polarize the light, they also act as waveplates. The orientation of the waveplate depends on the orientation of the plane of reflection. The retardance of the waveplate depends on the type of coating and on the angle of incidence. Data can be frightfully difficult to come by. It is generally assumed that dielectric coatings are worse than metal coatings. Silver is considerably better than aluminum.

The effect of these waveplates is to produce an optical path length through the system which depends on the orientation of the electric-field vector. If this polarization-dependent phase shift is different in the optical trains from the different telescopes to the beam combiner, the result is a loss of visibility. And the effect is large; even for silver coatings the phase shift can build up to a radian after only three or four reflections. Fortunately, the visibility loss depends on the difference in phase shift between the two beams being interfered. The

resulting design rule is to insist that the optical trains from each telescope to the beam combiner consist of the same number of reflections, each with the same angle of incidence and orientation and each with the same coating. If the coatings cannot be the same—for example with a reflection off a single beam splitter—that angle of incidence should be made as small as possible.

There is another reason to make the optical paths identical: beam rotation. Consider a 100% linearly polarized source and a feed system made from perfect mirrors. The light from each telescope at the beam combiner will also be linearly polarized. If the polarization vectors are not parallel, the visibility amplitude will be decreased by the cosine of that misalignment angle. Although it is harder to visualize, the same visibility loss occurs for unpolarized light.

The visibility loss due to the polarization-dependent phase shift can be solved by inserting a polarizer in the beam immediately before the detector. Because this *throws away half the light*, it may seem like a bad idea. But we need star light for different tasks—angle tracking and perhaps adaptive optics, fringe tracking, and science. Dividing the light by polarization for these different tasks may be an efficient way to handle the phase shift problems. However, the visibility loss due to beam rotation cannot be solved with a single polarizer.

14.3 Beam Combination and Modulation

Now we need to talk about beam combination and modulation. A single-baseline interferometer needs only to measure visibility amplitude, because the usefulness of a baseline phase is destroyed by the atmosphere. However, closure phases are needed to make an image and closure phases require three telescopes. To get the data needed for imaging, a beam combiner is needed that combines light from three or more telescopes simultaneously.

Given a number of telescopes, we need some means of combining the light from those telescopes and determining whether or not a fringe for each baseline is present in the data. If some or all of the fringes are missing, we need to move the delay lines to find them. Once a fringe is present, we have to calculate the amplitude and phase of that fringe. We will now discuss fringe-detection schemes and beam-combination schemes—topics that are closely related to each other.

14.3.1 Pupil-Plane Combination: Passive Detection

Tango and Twiss (1980) described what I will call *passive* detection. A simple two-way beam combiner for this method is illustrated in Figure 14.9. Light from two telescopes enter the beam combiner from the two inputs. They are combined at the surface of the beam splitter. The two combined beams are collected by two detectors. We label the intensity of the light detected by the two detectors A and B. If there is no fringe present

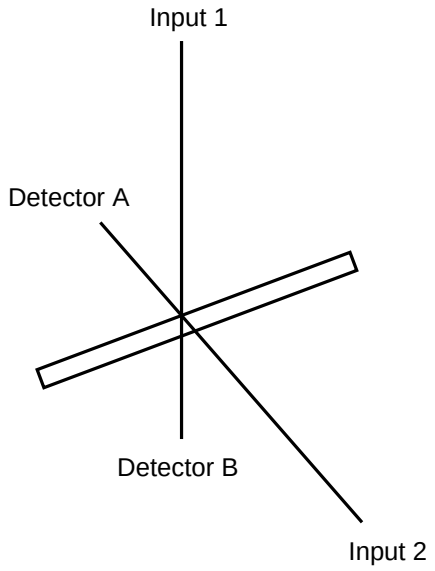


Figure 14.9: Beam combiner with two inputs and two outputs.

then $(A-B)$ will average to zero and have the same noise as $(A+B)$. If a fringe is present, then one side of the beam splitter will have a higher signal than the other side. Which side is higher depends on the phase of the fringe. As the fringe-phase varies, the fluctuations in $(A-B)$ will increase as the visibility amplitude increases but the fluctuations on $(A+B)$ will remain unchanged. The visibility amplitude can be determined by comparing the statistics of the sum and difference signals:

$$\langle V^2 \rangle = 2 \frac{\langle (A-B)^2 \rangle}{(\langle A \rangle + \langle B \rangle)^2}. \tag{14.2}$$

This is a very simple technique for measuring a fringe amplitude. A hidden assumption is that there has to be enough fringe motion during the observation for the fringe phase to be uniformly distributed. This should not be a problem most of the time but may become an issue for short observations, for high-precision measurements, or if a fringe tracker is used to partially stabilize the fringe packet. This is not a particularly useful method for imaging since phase information is discarded.

14.3.2 Image-Plane Combination: Spatial Modulation

As was discussed by Traub in Chapter 3, an image-plane beam combiner takes the beams of light from each of the telescopes and first lines them up so that they are parallel to each other, but separated by different spacings. This set of beams is passed through a single lens to form an image. For an unresolved star it is an Airy disk with fringes superimposed on it. The size of the Airy disk corresponds to the diameter of a single beam. The fringe frequencies depend on the spacing of beams. The two beams closest together will have the

widest fringes. The two beams furthest apart will have the highest frequency fringes. The phase of the fringe is the offset of the fringe peak from the center of the Airy disk. The easiest way to analyze the data is to take a Fourier transform. The transform will have a peak at each fringe frequency corresponding to a beam spacing. The amplitude and phase of the Fourier transform at a peak is the visibility amplitude and phase of the baseline formed from the two beams whose spacing at the lens corresponds to the fringe frequency. Needless to say, it is important to make sure no two beams have the same spacing.

14.3.3 Pupil-Plane Combination: Temporal Modulation

Pupil-plane beam combination brings the beams together on a beam splitter with zero beam spacing. Therefore the fringe frequency is also zero. As the phase changes, the image of the combined beam does not change shape as it does for the image-plane combination. The entire image changes brightness. Energy can still be conserved since images form on both sides of the beam splitter. As one image increases in brightness, the other becomes fainter.

To detect the fringe, the delay is modulated at a rate faster than the fringe motion induced by the atmosphere and the detectors are read out synchronously with that modulation. The signal varies sinusoidally with time. The phase of the signal relative to the phase of the modulation is the fringe phase.

Compared to image-plane detection, pupil-plane detection required fewer detectors but they have to be sampled faster.

Biases

The instrument and the atmosphere can bias the fringe measurement. The amplitude of the Fourier transform of the fringe data is not a delta function, it is a sinc function because of the finite length of the data stream. The peak occurs at the fringe frequency and it is the value at the peak that corresponds to the fringe visibility. If we do not know the exact ratio of the wavelength to the modulation length, the measured fringe amplitude will be wrong. And it will always be lower than the true value. You may think that this is not a problem; simply build an instrument with a stable modulation and measure it. It is not that simple. The effective wavelength is affected not only by the instrumental bandpass, but also by the star's spectrum and visibility variation across the bandpass. Fortunately, these effects are only important for the widest bandpass observations and the scientific interpretation of the result usually requires a narrower bandpass.

The atmosphere is a more important limitation since it is moving the fringe while the instrument is modulating the delay. As a result, the actual fringe frequency is a little higher or lower than the instrumental fringe frequency. Near the peak, the sinc function is quadratic. The error increases as the square of the atmospheric motion. This change

in fringe frequency is one of the few atmospheric calculations that is relatively easy. The atmospheric coherence time, t_0, is defined as the amount of time it takes the atmosphere to change the phase of the fringe by 1 radian. Therefore, the change in fringe frequency is

$$\delta f = \frac{1}{2\pi t_0} \tag{14.3}$$

On average, the fringe frequency differs from the instrumental fringe frequency by δf. The visibility amplitude error depends on the fractional fringe modulation error and can be decreased by increasing the modulation frequency. For typical atmosphere and observations in the visible, the modulation frequency should be on the order of 500 Hz for 1 percent measurements.

14.3.4 Demodulating Multiple Baselines

When light from multiple telescopes is combined on a single detector, the signals corresponding to the different baselines have to be separated. This is best done by imposing a different modulation on each baseline. The delay modulation has to be imposed on the light from each telescope before combination. Interpretation of the data is easiest and the resulting signal-to-noise is highest if the modulation uses a piecewise linear waveform: linear so that the modulation on the detector varies sinusoidally and can be analyzed using Fourier transforms, and piecewise so that the maximum delay can remain bounded. All of the delay modulations should be of the same frequency but of different amplitudes. The fringe frequency on a baseline is the difference in the modulation *amplitudes* of the two stations forming that baseline. If the modulation time is T and the difference in modulation amplitudes between the two stations forming the baseline is k wavelengths, then the fringe frequency is $f = k/T$. The modulation amplitudes must be chosen so that all of the fringe frequencies are different.

A example is shown in Table 14.1. This example uses six stations. There are 15 baselines and with modulation amplitudes of 0, 1, 4, 10, 12, and 17. There are no duplicate fringe frequencies. The fringe frequencies range from 1 to 17 with 14 and 15 missing. For more than four stations, it is always necessary to leave some intermediate frequencies unpopulated. These unpopulated frequencies can be useful for calibration.

The data from each linear modulation segment is Fourier transformed. The amplitude will have a peak at each modulation frequency. The amplitude and phase of that peak is the amplitude and phase of the fringe on the baseline corresponding to that fringe frequency.

Cross-Talk

An example of a multi-baseline data set can is shown in Figure 14.10. The upper panel shows the raw fringe data for a six-telescope, all-on-one combiner. It is a time series consisting of 15 sine waves of various amplitudes and phases. In this example, only the

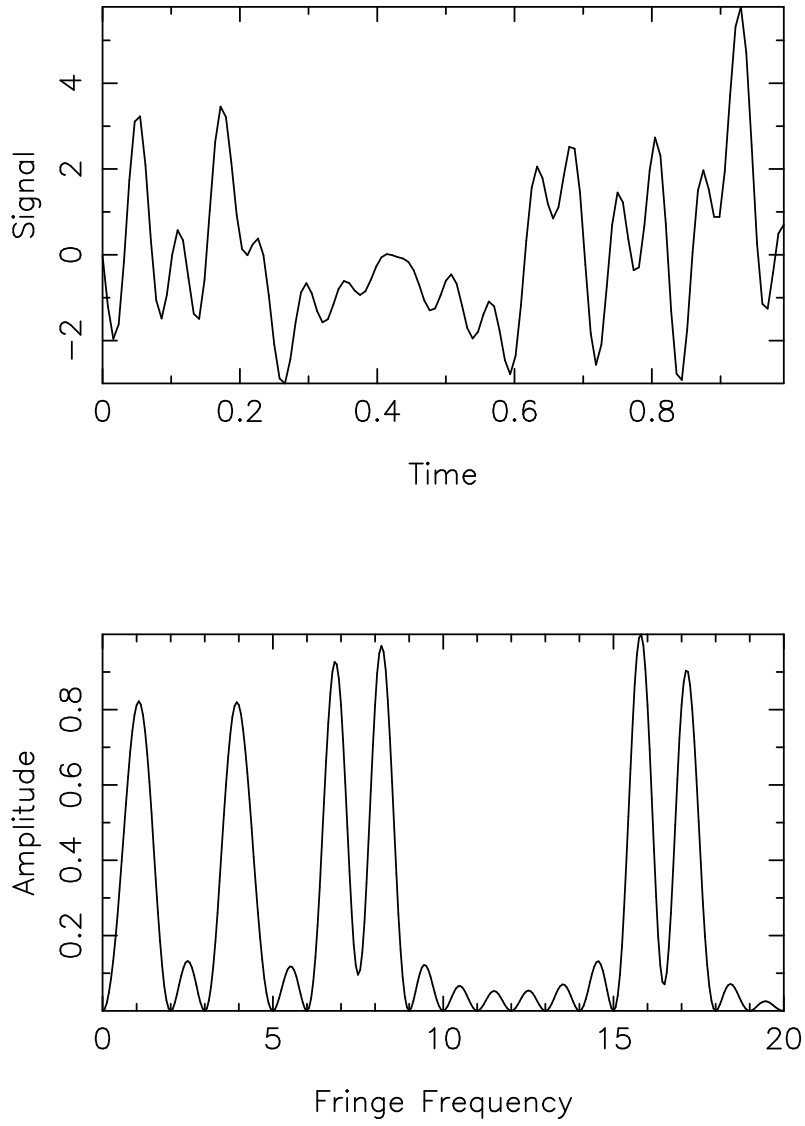


Figure 14.10: An example of multi-baseline data for six telescopes and 15 baselines. The upper panel shows the raw data as a time series. The lower panel shows the power spectrum. The height of a peak is proportional to the square of the visibility amplitude on that baseline.

Table 14.1: Demodulation schemes for multiple baselines.

Telescope Number	Modulation Amplitude	Baseline	Fringe Frequency
A	0		
B	1	A-B	1
C	4	A-C	4
D	10	A-D	10
E	12	A-E	12
F	17	A-F	17
		B-C	3
		B-D	9
		B-E	11
		B-F	16
		C-D	6
		C-E	8
		C-F	13
		D-E	2
		D-F	7
		E-F	5

six shortest baselines have high enough visibility amplitude for the fringes to be seen. The phases are random. The lower panel shows the power spectrum. The height of a peak is proportional to the square of the visibility amplitude on that baseline. Each of the peaks is a sinc function. The wiggles at frequencies between the six main peaks are sidelobes of those peaks. If the fringe frequencies are all integral multiples of the lowest fringe frequency, then each peak falls at a zero of all the other peaks and the measurements are independent. But the atmosphere moves the fringe frequencies around so we cannot maintain this condition and there is cross-talk between the baselines.

Even though this cross-talk is small, it is important. Remember that to make an image of a stellar surface, we need to measure visibility amplitudes on some baselines that are factors of 10 or more smaller than the tracking baselines. The cross-talk increases linearly with δf. As with the bias, the cross-talk can be decreased by increasing the modulation frequency, but for similar fringe parameter fidelity, it constrains the fringe frequency to be an order of magnitude higher than the constraint imposed by the bias.

One solution to this problem is to multiply the data by a tapered window function before Fourier transforming. This apodization reduced the height of the sidelobes at the price of increasing the width of the central peak. The result is a decrease in modulation frequency, but an increase in modulation amplitude—and a nearly complete elimination of cross-talk.

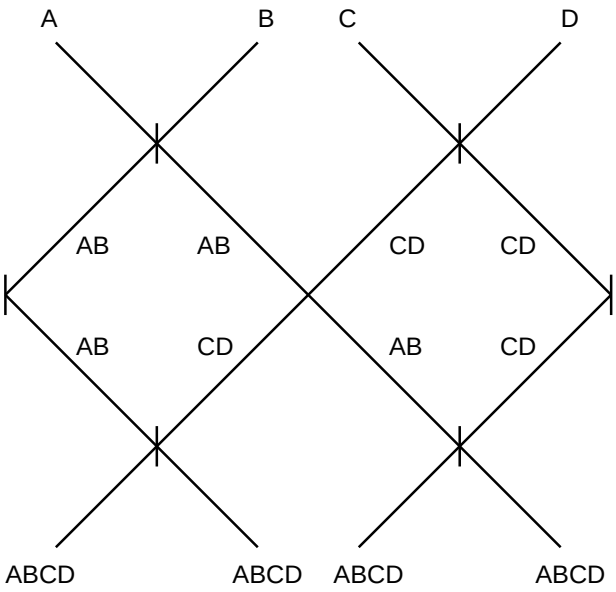


Figure 14.11: Beam combiner with four inputs and four outputs, similar to the one used at the Cambridge Optical Aperture Synthesis Telescope (COAST).

14.4 Beam Combination Techniques

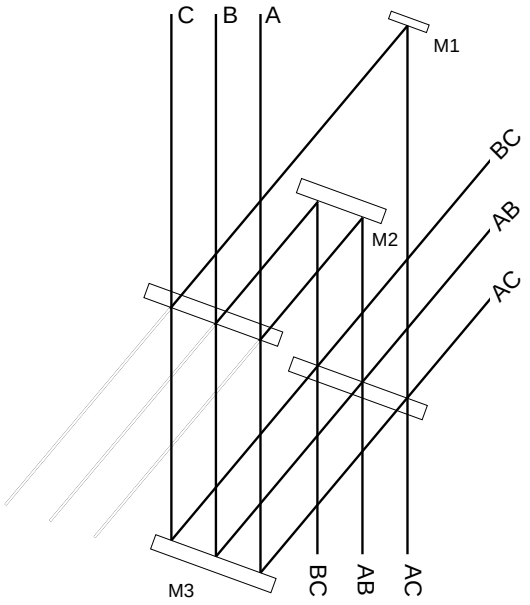
There is more to designing a beam combiner than deciding whether it will operate in the pupil plane or in the image plane. Beam topology plays an important role both in determining performance and in imposing performance requirements on the rest of the interferometer. Four possible topologies will be described and compared in this section.

14.4.1 All-On-One Combination

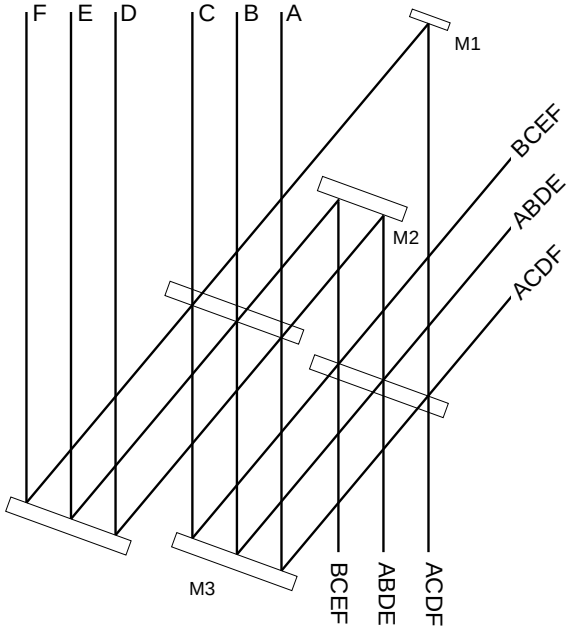
The beam combiner shown in Figure 14.11 is modeled after the one used at COAST, which is a multi-stage four-way beam combiner. Light from the telescopes is first combined in pairs. Then these pairs are recombined with other pairs. The result is that each detector sees light from *all* of the telescopes. This topology is easily extended to any power of two; combining light from 2^E telescopes requires 2^E detectors. A combiner designed for 2^E beams can be used for fewer beams by simply not using some of the inputs. It will still need 2^E detectors to collect all of the light.

14.4.2 Pairwise Combination

All of the beams do not have to be combined on a single detector. We could design a beam combiner that uses a different detector for each baseline. The example shown in the upper portion of Figure 14.12 was built at the NPOI. One advantage of this design over an all-on-



(a)



(b)

Figure 14.12: The Navy Prototype Optical Interferometer (NPOI) beam combiners. (a) a pairwise combiner for three beams. (b) An extension of the combiner in (a) for use with six beams. This hybrid design is intermediate in performance between a pairwise and an all-on-one combiner.

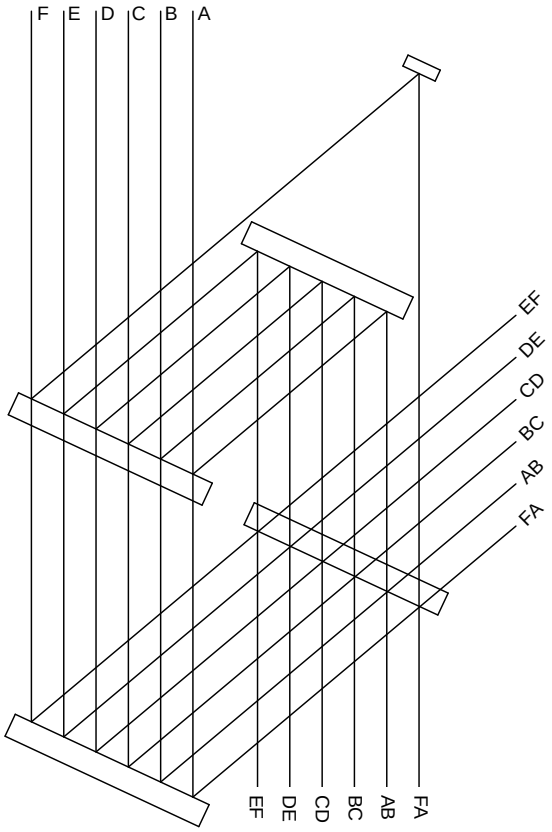


Figure 14.13: A partial pairwise beam combiner for six telescopes.

one design is that there is no chance of cross-talk. A disadvantage is that motion of any of the optics shown introduces a closure-phase error. A disadvantage of a pairwise combiner is that its complexity, and the number of detectors required, increases as the square of the number of beams, not linearly as in an all-on-one combiner.

The lower panel of Figure 14.12 shows an intermediate scheme that was adopted for six-way combination at the NPOI. The first beam splitter produces pairs which are passed through the three-way combiner to form quadruples. This topology requires fewer detectors and smaller modulation amplitudes than either pairwise or all-on-one combination.

14.4.3 Partial Pairwise Combination

A topology I am starting to like is the partial pairwise beam combiner. An example is shown in Figure 14.13 for six telescopes. There are six pairwise outputs corresponding to baselines A–B, B–C, C–D, D–E, E–F, and F–A. This is only 6 of the 15 possible baselines. As a result, this is not a particularly good science beam combiner, but it has enough functionality for a co-phasing beam combiner since fringe tracking on only five baselines is needed to stabilize the fringes on all of the baselines. Separating instrument functionality

into independent subsystems is a good thing to do from an engineering standpoint. If we are going to dedicate some of the star light to co-phasing the instrument, then this is a good topology for that job. But we need to worry since some light is already going to angle tracking. If we take some more for co-phasing, that leaves very little for science.

14.5 Comparison of Beam Combination Techniques

Many of the tradeoffs between different approaches to beam combination are difficult to quantify. For example how difficult it is to manufacture or maintain can depend on the person doing the work. The susceptibility of the data products to biases can depend on the types of observations planned for the instrument. As a result, this section ignores the truly important differences between beam combiners and focus on the one aspect easiest to quantify, sensitivity. We also ignore important details such as read noise and note that for photon-noise-limited performance, the signal-to-noise of an observation is a monotonically increasing a function of NV^2, where N is the number of photons in a single observation and V is the observed visibility amplitude. This is reasonable in the high signal-to-noise case.

$$\frac{\text{SIGNAL}}{\text{NOISE}} = \frac{NV}{\sqrt{N}} = \sqrt{NV^2}. \tag{14.4}$$

With this background, comparing the sensitivity of two beam combiners consists of calculating NV^2 for each configuration. The highest value wins. In everything that follows, N is the number of photons that reach a detector, N_0 is the number of photons from a single aperture, V is the visibility of a fringe at the detector, V_0 is the visibility that we would see if there were only two telescopes, and E is the number of telescopes contributing to the beam combiner.

Pairwise Combination

For pairwise combination, the photons from E telescopes are distributed between $E(E-1)/2$ detectors.

$$\begin{aligned} N &= \frac{EN_0}{E(E-1)/2} \\ &= 2\frac{N_0}{E-1}. \end{aligned} \tag{14.5}$$

The detector only sees light from the two telescopes contributing to the baseline so the visibility is not reduced. Therefore

$$NV^2 = (2N_0)V_0^2\left(\frac{1}{E-1}\right). \tag{14.6}$$

All-on-One Combination

For all-on-one combination, all of the detectors see the same signal. Since there is only photon noise in this example, the signals from all the detectors can be combined before being processed. The E telescopes contribute to a single detection, so $N = EN_0$.

The visibility is reduced because the fringe is formed between the light from two telescopes. As far as this baseline is concerned, the light from the other telescopes does not contribute to the fringe, it just forms a background against which the fringe detection has to be made. The visibility amplitude is reduced to $V = 2V_0/E$ and

$$NV^2 = (2N_0)V_0^2 \left(\frac{2}{E}\right). \tag{14.7}$$

since $2/E$ is larger than $1/(E-1)$, all-on-one combination is more sensitive than pairwise combination. Though this may seem like an odd result, it is telling us that the decrease in visibility amplitude due to adding in all those extra photons from the unused telescopes is more than offset by being able to use *all* the light from each telescope for each baseline. The gain is modest. The sensitivity advantage of an all-on-one combiner is 0.3 magnitudes for three telescopes, increasing with the number of telescopes to an asymptotic limit of 0.7 magnitudes.

CHARA: Partial-Pairwise Combination

The CHARA Array advocates the use a partial pairwise scheme (ten Brummelaar and Bagnuolo, 1994). Since, in general, this combiner will be used to co-phase an array, assume that only a fraction f of the light from each telescopes is used for beam combination. There are E telescopes and E detectors giving $N = fN_0$. Since this is a pairwise combiner, the visibility amplitude is not reduced and

$$NV^2 = (2N_0)V_0^2 \left(\frac{f}{2}\right). \tag{14.8}$$

The major advantage of this approach is now apparent; the sensitivity does not drop as more telescopes are added. The disadvantage is that it should be used in combination with a science fringe detector that detects all the baselines. As more photons are taken for the co-phasing (partial-pairwise) beam combiner, fewer photons reach the science beam combiner and the sensitivity is increased at the expense of pushing the science detectors into the photon-starved regime. But even with very few photons per integration for science, the phase of the fringe can be determined from the co-phasing beam combiner and observations can be combined to force the science data back into the photon-rich regime. This is clearly an interesting way to build an instrument.

Acknowledgments

The preparation of this talk made use of NASA's Astrophysics Data System Bibliographic Services.

References

J.T. Armstrong, D. Mozurkewich, L.J. Rickard, D.J. Hutter, J.A. Benson, P.F. Bowers, N.M. Elias II, C.A. Hummel, K.J. Johnston, D.F. Buscher, J.H. Clark III, L. Ha, L.-C. Ling, N.M. White, and R.S. Simon, "The Navy Prototype Optical Interferometer (NPOI)," Astrophys. J. **496**, 550–571 (1998).

T.A. ten Brummelaar and W.G. Bagnuolo, Jr., "The CHARA beam combiner design", in *Amplitude and Intensity Spatial Interferometry II*, J.B. Breckinridge, ed., Proc. SPIE **2200**, 140–151 (1994).

Y.L. Chow, "On designing a supersynthesis antenna array," IEEE Trans. Antennas Propagat. **20**, 30–35 (1972).

T.J. Cornwell, "A novel principle for optimization of the instantaneous Fourier plane coverage of correlation arrays," IEEE Trans. Antennas Propagat. **36**, 1165–1167 (1988).

M.J.E. Golay, "Point arrays having compact, nonredundant autocorrelations," J. Opt. Soc. Am. **61**, 272–273 (1970).

E. Keto, "The shapes of cross-correlation interferometers," Astrophys. J. **475**, 843–852 (1997).

E.M. Kong, D.W. Miller, R.J. Sedwick, "Optimal Trajectories and Orbit Design for Separated Spacecraft Interferometry," SERC No. 13-98 (November 1998).

N.C. Mathur, "A pseudodynamic programming technique for the design of correlator supersynthesis arrays," Radio Sci. **4**, 235–244 (1969).

D. Mozurkewich, "Hybrid design for a six way beam combiner," in *Amplitude and Intensity Spatial Interferometry II*, J.B. Breckinridge, ed., Proc. SPIE **2200**, 76–80 (1994).

A. Quirrenbach, D. Mozurkewich, D.F. Buscher, C.A. Hummel, and J.T. Armstrong, "Angular diameter and limb darkening of Arcturus," Astron. Astrophys. **312**, 160–166 (1996).

W.J. Tango and R.Q. Twiss, "Michelson stellar interferometry", Prog. Opt. **17**, 239–277 (1980).